

# On the efficiency of invariant multivariate sign and rank tests

KLAUS NORDHAUSEN   HANNU OJA   DAVID E. TYLER

**Abstract.** Invariant coordinate selection (ICS) is proposed in Oja and Tyler (2006) for constructing invariant multivariate sign and rank tests. The multivariate data vectors are first transformed to invariant coordinates, and univariate sign and rank tests are then applied to the components of the transformed vectors. In this paper, the powers of different versions of the one sample and two samples location tests are compared via simulation studies.

*2000 MSC codes:* 62H12, 62G10, 62G05.

*Key words and phrases:* Hodges Lehmann estimate; Kurtosis; M-estimate; Multivariate median; Transformation and retransformation technique; Wilcoxon test.

## 1 Introduction

The classical  $L_1$  type univariate sign and rank methods, estimates and tests, have been extended quite recently to the multivariate case. Multivariate extensions of the concepts of sign and rank based on (i) the vector of marginal medians, (ii) the so called spatial median or vector median, and (iii) the affine equivariant Oja median (Oja 1983) have been developed in a series of papers with natural analogues of one-sample, two-sample and multisample sign and rank tests. See e.g. Puri and Sen (1971), Möttönen and Oja (1995), Oja (1999), and Oja and Randles (2004) and references therein. These multivariate location estimates and tests are robust and nonparametric competitors of the classical MANOVA inference methods.

Unfortunately, the tests based on marginal signs and ranks and those based on spatial signs and ranks are not invariant under affine transformations of the observation vectors. Chakraborty and Chaudhuri (1996, 1998) and Chakraborty et al. (1998) introduced and discussed the so called transformation and retransformation technique to circumvent the problem: The data vectors are first linearly transformed back to a new, invariant coordinate system, the tests and estimates are constructed for these new vectors of variables, and, finally, the estimates are linearly retransformed to the original coordinate system. In the one sample and several samples  $p$ -variate location

problems, the transformation matrix was then based on  $p$  and  $p + 1$  original observation vectors, respectively.

Other nonparametric approaches for multivariate data analysis include the depth-based rank sum tests introduced by Liu and Singh (1993). The so called zonotopes and lift-zonotopes have been used to describe and investigate the properties of a multivariate distribution, see Mosler (2002). Randles (1989) developed an affine invariant sign test based on *interdirections*, and was followed by a series of papers introducing nonparametric sign and rank interdirection tests for multivariate one-sample and two-sample location problems. These tests are typically asymptotically equivalent with spatial sign and rank tests. Finally, in a series of papers, Hallin and Paindaveine constructed *optimal signed-rank tests* for the location and scatter problems in the elliptical model; see the seminal papers by Hallin and Paindaveine (2002, 2006).

In this paper, as proposed by Oja and Tyler (2006), two different scatter matrices are used to construct an invariant coordinate system. It is remarkable that, in the new coordinate system, the marginal variables are ordered according to their kurtosis. The multivariate variables are first transformed to invariant coordinates, and the univariate sign and rank tests are then applied to these transformed variables. Unlike most other invariant multivariate sign and rank methods, the resulting tests are distribution-free not only at elliptically symmetric models but rather at any symmetric model. The powers of different versions of the one sample and two samples location tests are compared via simulation studies.

Hence the structure of the paper is as follows. In Section 2 we introduce the basic notations and tools that are necessary to construct an invariant coordinate system and show its relationship with the kurtosis of the components. In Section 3 we point out different strategies to use univariate tests on the transformed data components to test the location problem in the one and two sample case. Section 4 gives results of a simulation study which compares the performance of the different strategies. The paper ends with a brief discussion in Section 5. For a complete discussion of this approach, see Oja and Tyler (2006).

## 2 Invariant coordinate selection (ICS)

### 2.1 Notations

Let  $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n$  be independent  $p$ -variate observations and write

$$Y = (\mathbf{y}_1 \ \mathbf{y}_2 \ \dots \ \mathbf{y}_n)$$

for the corresponding  $p \times n$  data matrix in the one sample case. In the several samples case, write

$$Y = (Y_1 \ \dots \ Y_c)$$

where  $Y_1, \dots, Y_c$  are independent random samples with sample sizes  $n_1, \dots, n_c$ ,  $n = n_1 + \dots + n_c$ , from  $p$ -variate distributions. In this paper we consider the one sample and two samples multivariate location problems only.

It is often desirable to have statistical methods which are invariant or equivariant under *affine transformations* of the data matrix, i.e. under transformations of the form

$$y_i \rightarrow Ay_i + \mathbf{b}, \quad i = 1, \dots, n,$$

or equivalently

$$Y \rightarrow AY + \mathbf{b}\mathbf{1}',$$

where  $A$  is a full-rank  $p \times p$  matrix and  $\mathbf{b}$  is a  $p$ -vector. The vector  $\mathbf{1}$  is a  $n$ -vector full of ones. Some interesting transformations are *orthogonal transformations* ( $Y \rightarrow UY$  with  $U'U = UU' = I$ ), *sign-change transformations* ( $Y \rightarrow JY$  where  $J$  is a  $p \times p$  diagonal matrix with diagonal elements  $\pm 1$ ), and *permutations* ( $Y \rightarrow PY$  where  $P$  is a  $p \times p$  permutation matrix obtained by successively permuting the rows and/or columns of  $I$ ). Note that transformation  $Y \rightarrow YP$  with a  $n \times n$  permutation matrix  $P$  permutes the observations.

## 2.2 Location vector and scatter matrices

We start by defining what we mean by a *location statistic*, a *scatter statistic*, and a *scatter statistic with respect to the origin*:

**Definition.** (i) A  $p$ -vector valued statistic  $T = T(Y)$  is called a *location statistic* if it is affine equivariant, that is,

$$T(A\mathbf{Y} + \mathbf{b}\mathbf{1}') = AT(\mathbf{Y}) + \mathbf{b}$$

for all full-rank  $p \times p$ -matrices  $A$  and for all  $p$ -vectors  $\mathbf{b}$ .

(ii) Second,  $p \times p$  matrix  $S = S(\mathbf{Y}) \geq 0$  is a *scatter statistic* if it is affine equivariant in the sense that

$$S(A\mathbf{Y} + \mathbf{b}\mathbf{1}') = AS(\mathbf{Y})A'$$

for all full-rank  $p \times p$ -matrices  $A$  and for all  $p$ -vectors  $\mathbf{b}$ .

(iii) Third, a *scatter statistic with respect to the origin* is affine equivariant in the sense that

$$S(A\mathbf{Y}J) = AS(\mathbf{Y})A'$$

for all full-rank  $p \times p$ -matrices  $A$  and for all  $n \times n$  sign change matrices  $J$ .

If  $\mathbf{Y}$  is a random sample, it is also natural to require that the statistics are invariant under permutations of the observations, that is,

$$T(YP) = T(\mathbf{Y}) \quad \text{and} \quad S(YP) = S(\mathbf{Y})$$

for all  $n \times n$  permutation matrices  $P$ .

In the semiparametric elliptic model, for example, the location statistic estimates the unknown center of symmetry  $\boldsymbol{\mu}$  and the scatter statistic  $\mathbf{S}(Y)$ , possibly multiplied by a correction factor, is an estimate of the regular covariance matrix  $\boldsymbol{\Sigma}$  if it exists. Different scatter statistics  $\mathbf{S}_1, \mathbf{S}_2, \dots$  then estimate the same population quantity but have different statistical properties (consistency, efficiency, robustness, computational convenience). In practice, one would choose the one that is most suitable for the problem at hand.

Different location and scatter statistics may also be used to construct skewness and kurtosis statistics; e.g. as in Kankainen et al. (2006),

$$\|\mathbf{T}_1 - \mathbf{T}_2\|_{\boldsymbol{\Sigma}}^2 \quad \text{and} \quad \|\mathbf{S}_1^{-1}\mathbf{S}_2 - \mathbf{I}\|^2$$

that is, the squared Mahalanobis distance between location statistics  $\mathbf{T}_1$  and  $\mathbf{T}_2$  and the squared matrix norm (Frobenius norm) of  $\mathbf{S}_1^{-1}\mathbf{S}_2 - \mathbf{I}$  where  $\mathbf{S}_1$  and  $\mathbf{S}_2$  (again equipped with correction factors) are different consistent estimates of the regular covariance matrix at the normal model. In this paper we will use two different scatter statistics to transform the data to invariant coordinates. See Section 2.4.

### 2.3 M-estimates of location and scatter

One of the earliest robust estimates developed for multivariate data are the M-estimates of multivariate location and scatter (Maronna 1976). The pseudo maximum likelihood (ML) estimates, including the regular mean vector and covariance matrix among others, are members of this class. Many other classes of estimates, like the S-estimates, CM-estimates and MM-estimates may be seen as special cases of M-estimates with auxiliary scale (Tyler 2002). M-estimates of location and scatter (one version),  $\mathbf{T} = \mathbf{T}(Y)$  and  $\mathbf{S} = \mathbf{S}(Y)$ , satisfy implicit equations

$$\mathbf{T} = [\text{ave}[w_1(r_i)]]^{-1} \text{ave}[w_1(r_i)\mathbf{y}_i]$$

and

$$\mathbf{S} = \text{ave}[w_2(r_i)(\mathbf{y}_i - \mathbf{T})(\mathbf{y}_i - \mathbf{T})']$$

for some suitably chosen weight functions  $w_1(r)$  and  $w_2(r)$ . The scalar  $r_i$  is the Mahalanobis distance between  $\mathbf{y}_i$  and  $\mathbf{T} = \mathbf{T}(Y)$ , that is,  $r_i = \|\mathbf{y}_i - \mathbf{T}\|_{\mathbf{S}}$ . Mean vector and covariance matrix are given by the choices  $w_1(r) = w_2(r) = 1$ .

If  $\mathbf{T}_1 = \mathbf{T}_1(Y)$  and  $\mathbf{S}_1 = \mathbf{S}_1(Y)$  are any affine equivariant location and scatter functionals then one-step M-functionals  $\mathbf{T}_2 = \mathbf{T}_2(Y)$  and  $\mathbf{S}_2 = \mathbf{S}_2(Y)$ , starting from  $\mathbf{T}_1$  and  $\mathbf{S}_1$ , are given by

$$\mathbf{T}_2 = [\text{ave}[w_1(r_i)]]^{-1} \text{ave}[w_1(r_i)\mathbf{y}_i]$$

and

$$\mathbf{S}_2 = \text{ave}[w_2(r_i)(\mathbf{y}_i - \mathbf{T}_1)(\mathbf{y}_i - \mathbf{T}_1)']$$

where now  $r_i = \|\mathbf{y}_i - \mathbf{T}_1\|_{\mathbf{S}_1}$ . It is easy to see that  $\mathbf{T}_2$  and  $\mathbf{S}_2$  are affine equivariant as well. Repeating this step until it converges yields a solution to the M-estimating equations with weight functions  $w_1$  and  $w_2$ . If  $\mathbf{T}_1$  is the mean vector and  $\mathbf{S}_1$  is the covariance matrix, then

$$\mathbf{T}_2 = \frac{1}{p} \text{ave}[r_i^2 \mathbf{y}_i] \quad \text{and} \quad \mathbf{S}_2 = \frac{1}{p+2} \text{ave}[r_i^2 (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})']$$

are one-step or reweighted M-estimates of location and scatter. Note that the scatter statistic  $\mathbf{S}_2 = \mathbf{S}_2(\mathbf{Y})$  is a *scatter matrix estimate based on fourth moments*. It is consistent for the regular covariance matrix at the multinormal model.

#### 2.4 Invariant coordinate selection

Scatter matrices are often used to standardize the data:

$$\mathbf{Y} \rightarrow \mathbf{Z} = [\mathbf{S}(\mathbf{Y})]^{-1/2} \mathbf{Y}.$$

Transformation matrix  $[\mathbf{S}(\mathbf{Y})]^{-1/2}$  thus yields the new coordinate system with uncorrelated components (in the sense of  $\mathbf{S}$ ). Unfortunately, this new coordinate system is not invariant under affine transformations; it is only true that

$$[\mathbf{S}(\mathbf{A}\mathbf{Y})]^{-1/2}(\mathbf{A}\mathbf{Y}) = \mathbf{U}[\mathbf{S}(\mathbf{Y})]^{-1/2} \mathbf{Y}$$

with an orthogonal matrix  $\mathbf{U}$  depending on  $\mathbf{Y}$ ,  $\mathbf{A}$  and  $\mathbf{S}$ .

Two different scatter functionals  $\mathbf{S}_1 = \mathbf{S}_1(\mathbf{Y})$  and  $\mathbf{S}_2 = \mathbf{S}_2(\mathbf{Y})$  may be used to find an invariant coordinate system as follows. For a more detailed discussion of the *invariant coordinate selection (ICS)*, see Oja and Tyler (2006). Starting with  $\mathbf{S}_1$  and  $\mathbf{S}_2$ , define a  $p \times p$  transformation matrix  $\mathbf{B} = \mathbf{B}(\mathbf{Y})$  and a diagonal matrix  $\mathbf{D} = \mathbf{D}(\mathbf{Y})$  by

$$\mathbf{S}_2^{-1} \mathbf{S}_1 \mathbf{B}' = \mathbf{B}' \mathbf{D}$$

that is,  $\mathbf{B}$  gives the eigenvectors of  $\mathbf{S}_2^{-1} \mathbf{S}_1$ . The following result can then be shown to hold.

**Result 1.** The transformation  $\mathbf{Y} \rightarrow \mathbf{Z} = \mathbf{B}(\mathbf{Y})\mathbf{Y}$  yields an *invariant coordinate system* in the sense that

$$\mathbf{B}(\mathbf{A}\mathbf{Y})(\mathbf{A}\mathbf{Y}) = \mathbf{J}\mathbf{B}(\mathbf{Y})\mathbf{Y}$$

for some  $p \times p$  sign change matrix  $\mathbf{J}$ . Matrix  $\mathbf{B}$  can be made unique by requiring that the element with largest absolute value in each row of  $\mathbf{B}$  is positive.

### 2.5 Kurtosis and ICS

Let  $\mathbf{B} = \mathbf{B}(\mathbf{Y})$  be the transformation matrix yielded by  $\mathbf{S}_1$  and  $\mathbf{S}_2$ . Observe that the elements of  $\mathbf{Z} = \mathbf{B}(\mathbf{Y})\mathbf{Y}$  are now standardized with respect to  $\mathbf{S}_1$  and uncorrelated with respect to  $\mathbf{S}_2$ , that is,

$$\mathbf{S}_1(\mathbf{Z}) = \mathbf{I} \quad \text{and} \quad \mathbf{S}_2(\mathbf{Z}) = \mathbf{D}$$

where  $\mathbf{D}$  is a diagonal matrix. The diagonal elements of  $\mathbf{D}$  yield the kurtosis measures for the components. Therefore the components of  $\mathbf{Z}$  are *ordered with respect to kurtosis*. Recall the discussion on kurtosis in Section 2.3.

In the simulations in this paper we use the invariant coordinate selection based on the regular covariance matrix  $\mathbf{S}_1$  and the scatter matrix  $\mathbf{S}_2$  based on the fourth moments. The  $j$ th diagonal element of matrix  $\mathbf{D}$  is then

$$D_{jj} = \frac{1}{p+2} \text{ave}_i \{z_{ij}^2(z_{i1}^2 + \dots + z_{ip}^2)\}, \quad j = 1, \dots, p.$$

Consider the case having some special interest in our simulations: Assume that  $\mathbf{Y} = \{\mathbf{y}_1 \dots \mathbf{y}_n\}$  is a random sample from a distribution which is a mixture of two multivariate normal distribution differing only in location:  $\mathbf{y}_i$  has a  $N_p(\mathbf{0}, \mathbf{I})$ -distribution with probability  $1 - \varepsilon$  and a  $N_p(\Delta\mathbf{e}_p, \mathbf{I})$ -distribution with probability  $\varepsilon$  ( $\varepsilon \leq 0.5$ ). (The last element in vector  $\mathbf{e}_p$  is one, other elements are zero.) Then  $\mathbf{S}_1(\mathbf{Y}) \rightarrow_p \mathbf{I}$  and  $\mathbf{S}_2(\mathbf{Y}) \rightarrow_p \mathbf{D}$  where  $\mathbf{D}$  is a diagonal matrix with  $D_{11} = \dots = D_{p-1,p-1} = 1$ . The last diagonal element is  $1 + b_2/(p+2)$  where  $b_2$  is the *classical univariate kurtosis* measure for the last component. Note that the last component has the highest kurtosis for  $\varepsilon < (3 + \sqrt{3})^{-1}$  and lowest kurtosis otherwise (compare Preston 1953). Also the amount of kurtosis strongly depends on the value of  $\Delta$ ; the greater  $\Delta$  the larger is the absolute value of kurtosis. This behavior is visualized in Figures 1 and 2.

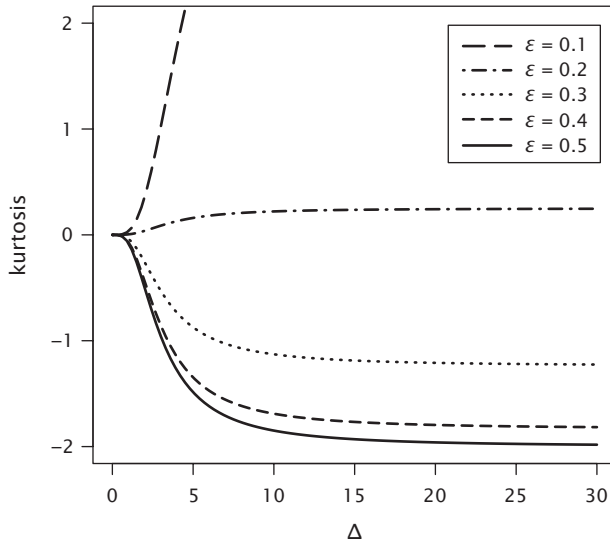
## 3 Invariant sign and rank tests

### 3.1 Marginal signs and ranks

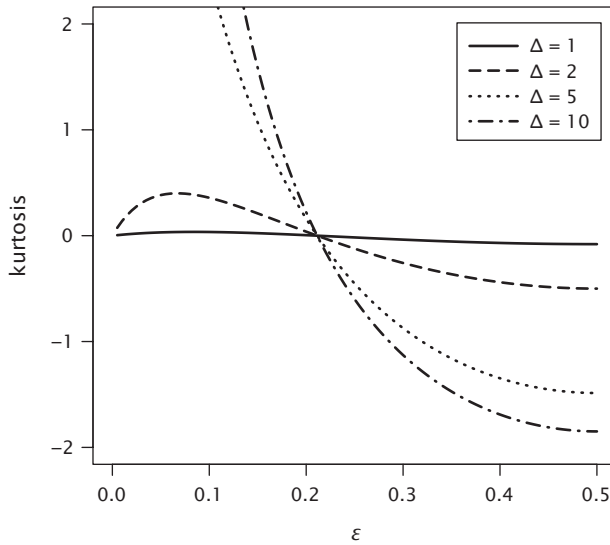
Let  $\mathbf{z}_i, i = 1, \dots, n$ , be the  $p$ -variate residuals in the multivariate location case, and consider the  $L_1$  type criterion functions

$$\text{ave}_i \{|z_{i1}| + \dots + |z_{ip}|\} \quad \text{and} \quad \text{ave}_{i,j} \{|z_{i1} - z_{j1}| + \dots + |z_{ip} - z_{jp}|\}.$$

The resulting  $L_1$  estimates are the vectors of marginal medians and marginal Hodges-Lehmann estimates. The corresponding score tests are based on the vectors of marginal (univariate) signs or marginal (univariate) ranks. See Puri and Sen (1971) for a complete discussion of this approach. The inference methods are invariant/equivariant under componentwise rescaling



**Figure 1.** Kurtosis for a location mixture of normal distributions as a function of  $\Delta$  for different  $\epsilon$ .



**Figure 2.** Kurtosis for a location mixture of normal distributions as a function of  $\epsilon$  for different  $\Delta$ .

but not orthogonally invariant/equivariant. The efficiencies do not exceed the univariate efficiencies and are quite low if the margins are highly correlated.

Invariant test versions can be obtained by first transforming the data to invariant coordinates. The use of the standardized data set  $[S(Y)]^{-1/2}Y$  does not help as the standardization is not affine invariant. See Section 2.4. Chakraborty and Chaudhuri (1996, 1998) avoided the problem by using  $p$  observations with indices listed in  $\alpha = (i_1, \dots, i_p)$ ,  $1 \leq i_1 < \dots < i_p \leq n$ , to construct, in the one-sample location case, a transformation matrix  $B(\alpha) = (\mathbf{y}_{i_1} \ \mathbf{y}_{i_2} \ \dots \ \mathbf{y}_{i_p})^{-1}$ . Now clearly  $B(\alpha)Y$  is invariant under affine transformations  $Y \rightarrow AY$  and the data set  $B(\alpha)Y$  may then be used for invariant one-sample test construction. In the several sample case, they choose  $\alpha = (i_1, \dots, i_{p+1})$ ,  $1 \leq i_1 < \dots < i_{p+1} \leq n$  and  $B(\alpha) = (\mathbf{y}_{i_1} - \mathbf{y}_{i_{p+1}} \ \mathbf{y}_{i_2} - \mathbf{y}_{i_{p+1}} \ \dots \ \mathbf{y}_{i_p} - \mathbf{y}_{i_{p+1}})^{-1}$ . This technique is then called the *transformation and re-transformation (TR) technique*. The problem naturally is how to choose  $\alpha$ , that is, the coordinate system in an optimal adaptive way. Techniques proposed for choosing  $\alpha$  tend to be computationally intensive since they require optimizing some criterion over all possible subsets of size  $p + 1$  from the sample. In the following we use the computationally simple invariant coordinate selection method based on two scatter matrices  $S_1$  and  $S_2$ .

### 3.2 One sample case

Let  $Y = (\mathbf{y}_1 \ \dots \ \mathbf{y}_n)$  be a random sample from a  $p$ -variate continuous distribution symmetric around unknown  $\boldsymbol{\mu}$ . We wish to test the null hypothesis  $H_0: \boldsymbol{\mu} = \mathbf{0}$  and estimate the unknown  $\boldsymbol{\mu}$ . For the test, let  $S_1$  and  $S_2$  be two scatter matrices with respect to the origin. Assume also that they are invariant under permutations to the observations. Then, for  $k = 1, 2$ ,

$$S_k(AYPJ) = AS_k(Y)A', \quad \forall A, P, J,$$

and therefore

$$B(YJP) = B(Y)$$

As, under the null hypothesis,  $Y$  is a random sample from distribution symmetric around the origin, it is also true that

$$Z(Y) \sim Z(Y)JP, \quad \forall J, P.$$

Clearly  $Z = (\mathbf{z}_1 \ \dots \ \mathbf{z}_n)$  is not a random sample any more. However, under the null hypothesis, the variables in  $(\mathbf{z}_1, \dots, \mathbf{z}_n)$  are exchangeable.

Consider next the  $j$ th component of the  $\mathbf{z}_i$  vectors, that is, the observations  $(z_{j1}, \dots, z_{jn})$ . Then, it is easy to see that

**Result 2.** Under the null hypothesis, the univariate *sign test statistic*

$$U_j = \sum_{i=1}^n I(z_{ji} > 0) \sim \text{Bin}(n, 0.5).$$

Thus, for all  $j = 1, \dots, p$ ,  $U_j$  is an invariant distribution-free multivariate sign test statistic. Unfortunately, the  $p$  sign test statistics  $U_1, \dots, U_p$  are not mutually independent.

Let next  $R_{ji}^+$  be the rank of  $|z_{ji}|$  among  $|z_{j1}|, \dots, |z_{jn}|$ . The univariate Wilcoxon signed-rank test statistic

$$W_j = \sum_{i=1}^n \operatorname{sgn}(z_{ji}) R_{ji}^+$$

is then distribution-free as well:

**Result 3.** Under the null hypothesis, the distribution of  $W_j$  is that of the one-sample Wilcoxon signed-rank test statistic.

The result easily follows from the facts that  $\operatorname{sgn}(z_{j1}), \dots, \operatorname{sgn}(z_{jn})$  are iid and independent of  $(|z_{j1}|, \dots, |z_{jn}|)$ . Also,  $|z_{j1}|, \dots, |z_{jn}|$  are exchangeable.

All the test statistics  $U_1, \dots, U_p$  and  $W_1, \dots, W_p$  are thus distribution-free but dependent (the dependence structure depends on the background distribution). How then to choose  $U_j$  or  $W_j$ , or how to combine these statistics for the testing problem? One goal of the present paper then is to provide some insight into this rather complex question. As the components are ordered according to their kurtosis, and one expects to see a high absolute value of kurtosis in the direction of  $\boldsymbol{\mu}$ , often the last (or first) component is most powerful and contains the most information. This fact can be utilised when constructing the “overall” test statistic where one can choose between different strategies. For example one could use only the first or only the last component or those two components combined. One could also use a rule like use the  $k \leq p$  components with the highest absolute value of kurtosis or one could simply use all components.

The corresponding *affine equivariant location estimates* are obtained as follows: Let  $\mathbf{T}$  be the vector of marginal medians or the vector of marginal Hodges-Lehmann estimators. These estimates are not location statistics as they are not affine equivariant. Let  $\mathbf{B} = \mathbf{B}(\mathbf{Y})$  be the transformation based on two scatter matrix estimates. Then multivariate affine equivariant *transformation-retransformation median* and *Hodges-Lehmann estimate* are obtained as

$$\tilde{\mathbf{T}}(\mathbf{Y}) = \mathbf{B}^{-1} \mathbf{T}(\mathbf{B}\mathbf{Y})$$

### 3.3 Two samples case

Let  $\mathbf{Y} = (\mathbf{Y}_1 \ \mathbf{Y}_2)$  where  $\mathbf{Y}_1$  and  $\mathbf{Y}_2$  are independent random samples of sizes  $n_1$  and  $n_2$ ,  $n = n_1 + n_2$ , from  $p$ -variate continuous distributions with cumulative density functions  $F(\mathbf{y})$  and  $F(\mathbf{y} - \boldsymbol{\mu})$ , respectively. We wish to test the null hypothesis  $H_0: \boldsymbol{\mu} = \mathbf{0}$  and estimate the unknown location shift  $\boldsymbol{\mu}$ . Let  $\mathbf{S}_1 = \mathbf{S}_1(\mathbf{Y})$  and  $\mathbf{S}_2 = \mathbf{S}_2(\mathbf{Y})$  be two scatter matrices *calculated from*

the combined data set and invariant under permutations to the observations. This is to say that, for  $k = 1, 2$ ,

$$S_k((AY + b1')P) = AS_k(Y)A', \quad \forall A, b, P,$$

and  $B(YP) = B(Y)$ . Under the null hypothesis, the combined sample  $Y = (Y_1 Y_2)$  is a random sample of size  $n$ , and

$$Z(Y) \sim Z(Y)P, \quad \forall P.$$

Again,  $Z = (z_1 \dots z_n)$  is not a random sample but, under the null hypothesis, the variables in  $(z_1, \dots, z_n)$  are exchangeable.

Affine invariant distribution-free multivariate rank tests may be constructed as follows. Let now  $R_{ji}$  be the rank of  $z_{ji}$  among  $z_{j1}, \dots, z_{jn}$ . As  $z_1, \dots, z_n$  are exchangeable,

**Result 4.** Under the null hypothesis the distribution of the univariate Wilcoxon rank test statistic

$$W_j = \sum_{i=n_1+1}^n R_{ji}$$

is that of regular two samples Wilcoxon test statistic with sample sizes  $n_1$  and  $n_2$ .

General rank score test statistics  $\sum_{i=n_1+1}^n a(R_{ji})$  may be constructed as well. The two samples sign test statistic (Mood's test statistic) is given by the choice  $a(i) = 1(0)$  for  $i > (\leq)(n + 1)/2$ . All the test statistics  $W_1, \dots, W_p$  are thus distribution-free but unfortunately dependent (the dependence structure depends on the background distribution). The question of which of those test statistics to use for the decision making allows the same strategies as in the one sample case.

Corresponding affine equivariant multivariate shift estimates are obtained as follows: Let  $T$  be the vector of marginal difference of the medians (Mood's test) or the vector of marginal two-sample Hodges-Lehmann shift estimators (Wilcoxon test). These estimates are not affine equivariant. Let  $B = B(Y)$  be the transformation based on two scatter matrix estimates. Then multivariate affine equivariant transformation retransformation estimates are again obtained as

$$\hat{T}(Y) = B^{-1}T(BY)$$

### 4 Simulation results

As mentioned in Section 3.2 and 3.3, several strategies are available for the decision making. We performed a simulation study to compare the following strategies in the one and two sample case:

- (i) Using a componentwise sign test and signed rank test as described in Puri and Sen (1971) based on all  $p$  components, denoted as  $U[1:p]$ , respectively as  $W[1:p]$ .
- (ii) Using the same componentwise sign test and signed rank test as before but only to combine the first and last component, denoted as  $U[1,p]$ , respectively as  $W[1,p]$ .
- (iii) Using an exact sign test respectively a Wilcoxon signed rank test for the last component only, denoted as  $U[p]$ , respectively as  $W[p]$ .

for different sample sizes and underlying distributions. As a reference test also Hotelling's  $T^2$  for the original observations is included. We note that both the exact and asymptotic distributions for case (i) and (ii) are still open questions. To approximate their distributions we suggest using distributions analogous to the asymptotic distributions given by Puri and Sen (1971), and conjecture that these approximate distributions are asymptotically correct. A size simulation (not shown here) supports this conjecture.

All simulations are based on 5000 repetitions and were performed using R 2.2.0 (R Development Core Team 2005) at the level  $\alpha = 0.05$ . The critical values for the tests were based on the limiting null distributions.

Not shown in the following subsections are results for the strategy which uses only the component with the largest absolute value of the kurtosis since this strategy had in all settings in the one sample case always less power than strategy (iii) and in the two sample case it was less powerful or about equal when compared to strategy (iii).

#### 4.1 One sample case

In this simulation we obtained the ICS with respect to the origin as described in Section 2.5 for data coming from a normal distribution and  $t_3$  and  $t_{10}$  distributions for different dimensions and sample sizes.

A size simulation (not shown here) yielded for all tests the designated level except for  $U[p]$  which was always smaller than 0.05 due to the discreteness of the test statistic and for Hotelling's  $T^2$  for heavy tailed distributions and small sample sizes.

To compare the power of the different strategies the location parameter of the distributions were set to  $\boldsymbol{\mu}_0 = (\Delta, 0, \dots, 0)'$  and  $\Delta$  in such a way chosen, that given the dimension  $p$  and the sample size  $n$  the power of Hotelling's  $T^2$  is 0.5 under normality. This means

$$P[F(p, n - p, \delta) > F_\alpha(p, n - p)] = 0.5$$

where  $F(p, n - p, \delta)$  is a random variable having a noncentral  $F$  distribution with degrees of freedom  $p$  and  $n - p$  and noncentrality parameter  $\delta = \frac{1}{n}\Delta^2$  and  $F_\alpha(p, n - p)$  is the  $1 - \alpha$  quantile of  $F(p, n - p) = F(p, n - p, 0)$ . This gives in our case a range for  $\Delta$  from 0.159 to 0.471.

The simulation results provided in Table 1 show that there is a lot of information in the last component, however the power of the strategies increases with the number of components they are based on and strategy (iii) can therefore not compete with strategy (i). Especially the signed rank test  $W[1:p]$  can be seen as a serious competitor to Hotelling's  $T^2$  since it is almost as efficient as Hotelling's  $T^2$  under normality and more efficient for heavier tails.

**Table 1.** Simulated power in the one sample case in number of rejections per 1000 cases.

Dist.	$p$	$n$	$T^2$	sign tests			signed rank tests		
				$U[1:p]$	$U[1,p]$	$U[p]$	$W[1:p]$	$W[1,p]$	$W[p]$
normal	2	50	499	340	340	208	472	472	323
		200	502	333	333	220	479	479	327
	5	50	500	281	180	122	441	257	203
		200	508	319	197	137	472	283	213
	10	50	507	204	124	89	385	168	159
		200	503	288	140	104	458	195	152
$t_{10}$	2	50	415	317	317	194	417	417	309
		200	413	324	324	213	429	429	298
	5	50	405	256	180	138	387	235	211
		200	414	301	195	139	419	255	193
	10	50	427	191	124	92	334	166	158
		200	409	283	138	101	417	185	147
$t_3$	2	50	261	286	286	180	334	334	257
		200	221	281	281	193	334	334	249
	5	50	244	237	169	117	299	200	182
		200	215	267	177	129	315	205	168
	10	50	270	173	130	95	267	155	149
		200	213	246	131	105	313	153	135

### 4.2 Two samples case

The setup for the two sample simulations are of a similar fashion as in the one sample case. The size simulation (also not shown here) gave similar results as in the one sample case, namely that the size of  $U[p]$  was always smaller than 0.05 and also Hotelling's  $T^2$  was smaller for heavier tails when the sample size was small.

The difference of the population locations  $\mu_0 = (\Delta, 0, \dots, 0)'$  was set also in such a way that under normality Hotelling's  $T^2$  would achieve a power of

0.5. The corresponding value of  $\Delta$  can then be computed via

$$P[F(p, n - p - 1, \delta) > F_\alpha(p, n - p - 1)] = 0.5$$

where the noncentrality parameter  $\delta$  is given as  $\delta = \frac{n_1 n_2}{n_1 + n_2} \Delta^2$ . This gives a range for  $\Delta$  from 0.223 to 0.637.

Table 2 shows the results for the two sample power simulations where in two settings the two populations are of equal size and in one setting the mixture probability is  $\varepsilon = 0.2$  (compare Section 2.5).

The same conclusions as for the one sample case apply basically also for the two sample case except one surprising occurrence for the rank test  $W[1:p]$  where the power drops considerably when the dimension and the sample sizes of both populations are large.

**Table 2.** Simulated power in the two sample case in number of rejections per 1000 cases.

Dist.	$p$	$n_1$	$n_2$	$T^2$	sign tests			signed rank tests		
					$U[1:p]$	$U[1,p]$	$U[p]$	$W[1:p]$	$W[1,p]$	$W[p]$
normal	2	50	50	504	321	321	177	482	482	321
		200	50	494	326	326	205	477	477	332
		200	200	494	329	329	203	477	477	317
	5	50	50	504	307	201	117	475	278	210
		200	50	491	309	191	137	464	267	199
		200	200	507	316	203	130	482	292	211
	10	50	50	499	259	136	86	449	192	159
		200	50	484	282	144	99	443	198	154
		200	200	501	212	145	92	310	199	153
$t_{10}$	2	50	50	404	304	304	170	423	423	297
		200	50	405	310	310	214	418	418	307
		200	200	409	306	306	195	421	421	294
	5	50	50	402	277	182	114	409	252	207
		200	50	400	290	195	135	422	256	201
		200	200	393	290	191	121	405	255	194
	10	50	50	422	251	130	89	416	186	167
		200	50	414	293	142	97	421	179	146
		200	200	410	189	132	86	280	176	138
$t_3$	2	50	50	233	277	277	160	334	334	244
		200	50	219	278	278	179	330	330	239
		200	200	214	285	285	179	336	336	246
	5	50	50	233	249	177	102	320	221	175
		200	50	213	254	181	122	321	211	164
		200	200	194	268	174	110	318	210	152
	10	50	50	230	204	123	72	296	145	132
		200	50	209	241	136	99	306	152	126
		200	200	197	183	135	84	211	149	119

## 5 Final comments

This simulation study serves as an introduction to the use of two different scatter matrices to obtain an ICS where invariant sign and rank tests can be constructed. It is obvious that invariance of the test statistics is a worthwhile aim to pursue and the ICS is a promising tool to achieve this goal and has for example compared to the TR technique the advantage that not  $p$ , respectively  $p + 1$ , data points have to be singled out on which the transformation depends on. However for the ICS a choice of the two scatter matrices must be made and further research is necessary to compare the effect of different choices. For instance from a nonparametric point of view the assumption of fourth order moments as in this study is not fortunate. Also surprising for us was that contrary to the spatial sign test in the elliptical case for large  $n$  and  $p$  the efficiencies of the tests used here seem not to tend to 1 in the two sample case.

Another point to pursue would be the efficiencies of the tests for different values of  $\Delta$  which would occur for example if a larger power for Hotelling's  $T^2$  would be required because then, as can be seen in Figure 1, the main direction of the data would become more distinct given in the two sample case that the mixing probability  $\varepsilon$  would be not too close to  $1/(3 + \sqrt{3})$ .

## Acknowledgements

The work of Dave Tyler was supported by the NSF Grant DMS-0305858. The work of Klaus Nordhausen and Hannu Oja was supported by grants from Academy of Finland.

## References

- Chakraborty, B. and Chaudhuri, P. (1996). On a transformation retransformation technique for constructing affine equivariant multivariate median. *Proceedings of American Mathematical Society*, 124, 1529-1537.
- Chakraborty, B. and Chaudhuri, P. (1998). On an adaptive transformation retransformation affine equivariant estimate of multivariate location. *Journal of the Royal Statistical Society, Series B*, 60, 145-157.
- Chakraborty, B., Chaudhuri, P., and Oja, H. (1998). Operating transformation retransformation on spatial median and angle test. *Statistica Sinica*, 8, 767-784.
- Hallin, M. and Paindaveine, D. (2002). Optimal tests for multivariate location based on interdirections and pseudo-Mahalanobis ranks. *Annals of Statistics*, 30, 1103-1133.
- Hallin, M. and Paindaveine, D. (2006). Optimal rank-based tests for sphericity. *Annals of Statistics*, to appear.
- Kankainen, A., Taskinen, S., and Oja, H. (2006). Tests of multinormality based on location vectors and scatter matrices. Submitted.
- Liu, R. Y. and Singh, K. (1993). A quality index based on data depth and multivariate rank tests. *Journal of the American Statistical Association*, 88, 252-260.

- Maronna, R. A. (1976). Robust M-estimators of multivariate location and scatter. *Annals of Statistics*, 17, 1608-1630.
- Mosler, K. (2002). *Multivariate Dispersion, Central Regions and Depth: The Lift Zonoid Approach*. Lecture Notes in Statistics, Vol. 165. New York: Springer.
- Möttönen, J. and Oja, H. (1995). Multivariate spatial sign and rank methods. *Journal of Nonparametric Statistics*, 5, 201-213.
- Oja, H. (1983). Descriptive statistics for multivariate distributions. *Statistics & Probability Letters*, 1, 327-332.
- Oja, H. (1999). Affine invariant multivariate sign and rank tests and corresponding estimates: A review. *Scandinavian Journal of Statistics*, 26, 319-343.
- Oja, H. and Randles, R. (2004). Multivariate nonparametric tests. *Statistical Science*, 19, 598-605.
- Oja, H. and Tyler, D. E. (2006). Invariant multivariate sign and rank tests. *Manuscript in preparation*.
- Preston, E. J. (1953). A graphical method for the analysis of statistical distributions into two normal components. *Biometrika*, 40, 460-464.
- Puri, M. L. and Sen, P. K. (1971). *Nonparametric Methods in Multivariate Analysis*. New York: Wiley & Sons.
- R Development Core Team (2005). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Randles, R. H. (1989). A distribution-free multivariate sign test based on interdirections. *Journal of the American Statistical Association*, 84, 1045-1050.
- Tyler, D. E. (2002). High breakdown point multivariate M-estimation. *Estadística*, 52, 213-247.

KLAUS NORDHAUSEN  
Tampere School of Public Health  
FI-33014 University of Tampere, Finland  
Klaus.Nordhausen@uta.fi  
<http://www.uta.fi/~klaus.nordhausen/>

HANNU OJA  
Tampere School of Public Health  
FI-33014 University of Tampere, Finland  
Hannu.Oja@uta.fi  
<http://www.uta.fi/~hannu.oja/>

DAVID E. TYLER  
Department of Statistics  
The State University of New Jersey  
Piscataway NJ 08854, USA  
dtyler@rci.rutgers.edu  
<http://www.rci.rutgers.edu/~dtyler/>